Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Prediction of Real-World Drug Effectiveness Pre-Launch: Case Study in Rheumatoid Arthritis

Eva-Maria Didden, PhD[1], Yann Ruffieux, MSc[1], Noemi Hummel, PhD[1], Orestis Efthimiou, PhD[1,2], Stephan Reichenbach, MD[1,3], Sandro Gsteiger, PhD[4], Axel Finckh, MD[5] , Christine Fletcher, MSc[6], Georgia Salanti, PhD[1,2], Matthias Egger, MD[1*], on behalf of IMI GetReal Work Package 4

[1] Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

[2] University of Ioannina, School of Medicine, Ioannina, Greece

[3] Department of Rheumatology, Immunology and Allergology, University Hospital and University of Bern, Switzerland

[4] F. Hoffmann-La Roche Ltd, MORSE - Health Technology Assessment Group, Basel, Switzerland

[5] University Hospital of Geneva (HUG), Switzerland

[6] Amgen Ltd, Cambridge, Great Britain

Address correspondence to:

*Professor Matthias Egger, Institute of Social and Preventive Medicine (ISPM), Finkenhubelweg 11, CH-3012 Bern, Switzerland. Telephone: +41 31 631 35 01. Email: matthias.egger@ispm.unibe.ch

Funding: The work leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° [115546], resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007- 2013) and EFPIA companies' in kind contribution. The Swiss Clinical Quality Management in Rheumatic Diseases (SCQM) is supported by the pharmaceutical companies and donors: a list of financial supporters can be found on www.scqm.ch/sponsors. The collection of data in the British Society for Rheumatology Biologics Registry in

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

RA (BSRBR-RA) is funded via a partnership between the British Society of Rheumatology and pharmaceutical companies (currently Abbvie, MSD, Pfizer, Roche and UCB).

Key words: Efficacy-effectiveness gap, treatment predictor, prognostic factor, effect modifier, prediction model, rheumatoid arthritis.

Running title: Predicting drug effectiveness

Acknowledgements: The research leading to these results was done within the framework of Work Package 4 of the GetReal consortium. For further information, please refer to http://www.imi-getreal.eu/. This paper only reflects the personal views of the stated authors. We express our thanks for a fruitful collaboration to all GetReal Work Packages. Our sincere thanks go to F. Hoffmann La-Roche Ltd., especially to Aijing Shang and Máximo Carreras, to the Swiss Clinical Quality Management in Rheumatic Diseases (SCQM) and its contributing institutions (http://www.scqm.ch/institutions/), especially to Almut Scherer. The BSRBR-RA, a UK based cohort study which monitors the use and safety of biologic agents in rheumatological medical practice, provided data used in this study. The collection of data in BSRBR-RA is funded by restricted income via a partnership between the British Society of Rheumatology and pharmaceutical companies currently Abbvie, MSD, Pfizer, Roche and UCB.

Conflict of interest statement: The authors have no conflicts of interest to declare

Word count: Abstract 250 words, main text 4033 words, 3 tables, 2 figures, 1 box, 31 references, 1 document with supplemental materials.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Abstract

Background: Decision-makers often need to assess the real-world effectiveness of new drugs pre-launch, when phase II/III randomized controlled trials (RCTs) but no other data are available.

Objective: To develop a method to predict drug effectiveness pre-launch and to apply it in a case study in rheumatoid arthritis (RA).

Methods: The approach (1) identifies a market-approved treatment (S) currently used in a target population similar to that of the new drug (N); (2) quantifies the impact of treatment, prognostic factors and effect modifiers on clinical outcome; (3) determines the characteristics of patients likely to receive *N* in routine care; (4) predicts treatment outcome in simulated patients with these characteristics. Sources of evidence include expert opinion, RCT and observational studies. The framework relies on generalized linear models.

Results: The case study assessed the effectiveness of tocilizumab (TCZ), a biologic Disease-Modifying Anti-Rheumatic Drug (DMARD), combined with conventional DMARDs, compared to conventional DMARDs alone. Rituximab (RTX) combined with conventional DMARDs was identified as treatment *S*. Individual participant data from two RCTs and two national registries were analyzed. The model predicted the 6-months changes in the Disease Activity Score 28 (DAS28) accurately: the mean change was -2.101 (standard deviation (SD): 1.494) in the simulated patients receiving TCZ and conventional DMARDs as compared to -1.873 (SD: 1.220) in retrospectively assessed observational data. It was -0.792 (SD: 1.499) in registry patients treated with conventional DMARDs.

Conclusion: The approach performed well in the RA case study, but further work is required to better define its strengths and limitations.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Introduction

The question whether and to what extent estimates of the efficacy of drugs from randomized controlled trials (RCTs) will reflect the drugs' effectiveness in real-world routine care is of great importance for drug developers, regulators, reimbursement agencies and tax payers [1,2]. The potential difference between RCT outcomes and effects in real-life settings has been called the 'efficacy–effectiveness gap', which may be due to the greater variability of patients receiving the drug in real world compared to study settings [3,4]. The response to drugs differs across patients due to biological factors, for example depending on the expression of a receptor, or the presence or absence of an allele, or the severity of the underlying disease or co-morbidity [4,5]. Behavioral and socio-economic factors may also be important, for example poor adherence to prescribed schedules [4].

Comparisons of the efficacy of a drug observed in clinical trials and its effectiveness in routine care are typically performed when the drug is on the market and observational data on its effectiveness in the real world has accumulated [6]. However, for decision-makers, data on the likely real-world effectiveness of a new drug would be of particular interest before the drug enters the market. Studies predicting treatment effects not directly assessed in existing RCTs such as treatment effects in different (real-world) patient populations or settings, long-term outcomes, or different doses are rare. A recent systematic review identified only 12 such studies, which typically examined cardiovascular and metabolic diseases or neurological conditions using mathematical or statistical modelling [7]. Several of these studies have been widely cited, and mentioned in clinical guidelines. For example, the Seattle heart failure model [8] has been recommended in American and European guidelines on the management of heart failure [9,10] and the CDC diabetes cost-effectiveness model [11] in the Canadian Diabetes Association guidelines [12].

Within the European Union's Innovative Medicines Initiative "GetReal: Incorporating real-life data into drug development" [13,14] we developed a modelling framework to predict the real-world effectiveness of a new treatment at a point in time when this new treatment is still in the market approval process. We illustrate our approach with a case study of the treatment of rheumatoid arthritis (RA) with biologic and conventional, non-biologic Disease Modifying Anti-Rheumatic Drugs (DMARDs). The proposed prediction model may be valuable for stakeholders interested in the use of modelling and simulation approaches to support decision making, including outcome researchers and health economists in the pharmaceutical industry or regulatory and reimbursement agencies. The framework may also be helpful in post-launch health technology assessment (HTA) and guideline development when observational data exist for some, but not all, interventions.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Case study: the treatment of rheumatoid arthritis

Biologic and conventional, non-biologic DMARDs are often prescribed in combination for the treatment of RA [15]. Several widely used biologic DMARDs inhibit the Tumor Necrosis Factor (TNF), a substance that causes inflammation ("anti-TNF agents"). Rituximab (RTX) is also a biologic DMARD: a monoclonal antibody directed at the CD20 receptor of B lymphocyte cells, which depletes peripheral B cells. Methotrexate (MTX) is a commonly used conventional synthetic DMARD (cDMARD). In clinical practice, the choice of DMARD depends on several factors, including patient characteristics, disease history and previous medications.

In our hypothetical case study, we aimed to assess the real-world effectiveness prior to launch of tocilizumab (TCZ), a relatively new, cytokine-directed biologic DMARD. We assumed that TCZ is still in the market approval process and that observational evidence is available for other biologic DMARDs, but not for TCZ. We used individual participant data from the Tocilizumab in Combination With Traditional DMARD Therapy (TOWARD) and the Randomized Evaluation of Long-Term Efficacy of Rituximab (REFLEX) trials [16,17] and individual participant data from two national RA registries, the Swiss Clinical Quality Management in rheumatic diseases (SCQM) [18] and the British Society for Rheumatology Biologics Registry in RA (BSRBR-RA) [19]. The outcome of interest was the change in Disease Activity Score 28 (DAS28), calculated using the erythrocyte sedimentation rate, which is the clinically most universal disease severity index in RA [20,21]. A lower score indicates lower disease activity. We did no investigate other disease severity indices due to high rates of missing values in the observational databases. Further details on the trials and registries analyzed in this study are given in supplemental Box S1.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Methods

A. Data

Sources of data

We assume that RCT data are available comparing the new drug of interest, denoted by N, to control, C. Data from observational studies are available for C and a third treatment, S, which is considered to be prescribed to a target population similar to that anticipated for N. No observational evidence is available for *N*.

Variables

We assume that RCTs and observational studies collected and reported data in comparable ways, including laboratory, follow-up visit and outcome data, and that sample sizes are sufficiently large to provide reasonably precise estimates. Prognostic factors and effect modifiers are available in both data sources, and factors that are predictive for the use of *S* are available in the observational studies. Prognostic factors are covariates that affect the natural course of the disease, independently of treatment. Effect modification refers to the situation where the relative treatment effect depends on the value of at least one other covariate, the effect modifier. The four steps of our framework are described below and summarized in supplemental Box S2. Appendix S2 provides a detailed description of the suggested model selection process.

B. Statistical modelling and prediction framework

Identification of licensed drug currently used in similar patient populations

The first step consists of identifying an approved drug *S* that is interchangeable with the new drug *N* in terms of the characteristics of the patients who will be prescribed the drug. This step requires expert knowledge. We assume that a drug *S* has been identified. The observational data provide information on the profile of patients receiving the drug in routine practice and on the effectiveness of S. If RCT data on the efficacy of S are available, they may be included and their influence examined using the weighted approach described in Appendix S1.

Estimation of the impact of prognostic factors and assessment of efficacy of the new drug, accounting for effect modifiers

In a second step a generalized linear model is developed. The notation is given in Box 1; bold face highlights matrices or vectors. By y_i, we denote the outcome observed within a pre-specified timeframe in a participant i

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

who either receives treatment N or S ($T_i = 1$) or a comparator intervention C ($T_i = 0$) and whose prognostic factors and effect modifiers are captured in the vectors $\boldsymbol{x_i^{PF}}$ and $\boldsymbol{x_i^{EM}}$.

The expected treatment response $E(y_i)$ in patient i is then modelled using a generalized linear model

$$E(y_i) = \begin{cases} g\left(\beta_0 + \boldsymbol{\beta^{PF}} \boldsymbol{x_i^{PF}}\right) & \text{if patient i received } C \\ g\left(\beta_0 + \boldsymbol{\beta^{PF}} \boldsymbol{x_i^{PF}} + \mu_N + \boldsymbol{\beta_N^{EM}} \boldsymbol{x_i^{EM}}\right) & \text{if patient i received } N \\ g\left(\beta_0 + \boldsymbol{\beta^{PF}} \boldsymbol{x_i^{PF}} + \mu_S + \boldsymbol{\beta_S^{EM}} \boldsymbol{x_i^{EM}}\right) & \text{if patient i received } S \end{cases} \quad (1)$$

Suitable response functions $g(\cdot)$ depend on the nature of the outcome. The approach is completed with a sampling model for y_i, $y_i \sim F(E(y_i), \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ corresponds to a set of (potential) nuisance parameters estimated along with the other unknown model components. A conventional linear model, for instance, would imply that $y_i \sim N(E(y_i), \boldsymbol{\Psi} = \sigma^2)$, with σ^2quantifying the model variance. In this case, no transformation through $g(\cdot)$ is needed. Centering all continuous covariates around their population means, $g(\beta_0)$ can be interpreted as the expected treatment outcome in a patient taking C, who represents the "average" overall patient population. Choosing drug j rather than C for this patient, the relative change in the expected treatment outcome would be reflected through μ_j. We distinguish between the relative treatment effect of N estimated from RCT data (μ_N) and the relative effect of S estimated from observational data (μ_S). The row vector $\boldsymbol{\beta^{PF}}$contains parameters that quantify the impact of prognostic factors. The impact of effect modifiers in the RCT data (for *N*) can be obtained from row vector $\boldsymbol{\beta_N^{EM}}$, while the impact of effect modifiers in the observational data (for *S*) are captured by row vector $\boldsymbol{\beta_S^{EM}}$.

We fit Equation (1) to the observational and RCT data to estimate the unknown parameters and their variance-covariance matrix. Both RCT and observational evidence are thus considered in the estimation of the intercept term β_0 and the vector of the prognostic effects $\boldsymbol{\beta^{PF}}$. Since we do not expect *N* to have the same efficacy as *S*, the evidence about relative treatment effects and effect modification should come from RCT data alone. This is addressed in Equation (1) which separates RCT evidence on N from observational evidence on S. If individual participant data from more than one RCT or more than one observational study are available, the user may want to weigh the different databases according to their quality or relevance. In supplemental Appendix S1, we present an alternative modelling strategy, which uses weights.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Determination of profile of patients who are likely to receive the new treatment in the real world of a healthcare system

In a third step we assume that the likelihood of a patient being prescribed a drug *j* rather than starting or continuing with a comparator intervention C is driven by patient and disease characteristics as well as properties of the health care system. These factors are captured in the vector of treatment predictors $\boldsymbol{x_i^{TP}}$. For patient i, this is modelled with a logistic regression equation

$$logit(p(T_i = 1)) = \gamma_0 + \boldsymbol{\gamma x_i^{TP}}, \quad (2)$$

where the row vector $\boldsymbol{\gamma}$ quantifies the impact of the treatment predictors on the log-odds to receive drug *j* rather than *C*. Assuming, again, a centered covariate setting, the log-odds in the "average" patient population would be γ_0. Here, we only use observational data on S and C to estimate γ_0 and $\boldsymbol{\gamma}$, since we assume that such data are not available on N. Assuming that S and N are interchangeable in terms of treatment decision, the model of receiving *S* versus C can be used to predict receipt of *N* versus C.

Prediction of outcome in patients who will likely receive the new drug in the real world of a healthcare system

To predict outcomes in the patients likely to receive the new drug *N*, we simulate a population of n patients from a multivariate normal distribution. For each patient we generate a value for each prognostic factor, effect modifier, and treatment predictor. We construct the underlying distribution using the empirical means and covariances of the relevant variables from an observed sample that is representative of the target population. Categorical variables are treated as continuous variables in the simulation process, then transformed back into discrete variables [22]. Each simulated patient is denoted by index i^*, i.e. $i^* \in \{1, \dots, n\}$, and characterized by covariate vectors $\boldsymbol{x_{i*}^{PF}}$, $\boldsymbol{x_{i*}^{EM}}$, $\boldsymbol{x_{i*}^{TP}}$. We further consider the two sets of parameter estimates $\widehat{\boldsymbol{\Theta}}^{\boldsymbol{I}} \coloneqq \{\hat{\beta}_0, \hat{\mu}_N, \widehat{\boldsymbol{\beta}}^{\boldsymbol{PF}}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{N}}^{\boldsymbol{EM}}\}$ and $\widehat{\boldsymbol{\Theta}}^{\boldsymbol{II}} \coloneqq \{\hat{\gamma}_0, \widehat{\boldsymbol{\gamma}}\}$ and their variance-covariance matrices $\boldsymbol{V}(\widehat{\boldsymbol{\Theta}}^{\boldsymbol{I}})$ and $\boldsymbol{V}(\widehat{\boldsymbol{\Theta}}^{\boldsymbol{II}})$, as derived from Equations (1) and (2). To account for uncertainty in parameter estimation, each individual i^* is assigned one random draw $\widetilde{\boldsymbol{\Theta}}_{\mathbf{i*}}^{\boldsymbol{I}}$, where $\widetilde{\boldsymbol{\Theta}}_{\mathbf{i*}}^{\boldsymbol{I}} \sim N(\widehat{\boldsymbol{\Theta}}^{\boldsymbol{I}}, \boldsymbol{V}(\widehat{\boldsymbol{\Theta}}^{\boldsymbol{I}}))$, and another independent draw $\widetilde{\boldsymbol{\Theta}}_{\boldsymbol{i*}}^{\boldsymbol{II}}$, where $\widetilde{\boldsymbol{\Theta}}_{\boldsymbol{i*}}^{\boldsymbol{II}} \sim N(\widehat{\boldsymbol{\Theta}}^{\boldsymbol{II}}, \boldsymbol{V}(\widehat{\boldsymbol{\Theta}}^{\boldsymbol{II}}))$. We thus obtain a total of n samples from each of the two normal distributions. The predicted treatment decision for patient i^*, $\tilde{T}_{i*}$, is then sampled from a Bernoulli distribution

$$\tilde{T}_{i*} \sim Ber\left(expit\left(\tilde{\gamma}_{0,i*} + \widetilde{\boldsymbol{\gamma}}_{\boldsymbol{i*}}\boldsymbol{x_{i*}^{TP}}\right)\right)$$

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

where $expit(\cdot)$ is the inverse function of the log-odds function $logit(\cdot)$. Afterwards, the predicted treatment outcome of patient i^*, $\tilde{y}_{i*}$, is obtained from the sampling model

$$\tilde{y}_{i*} \sim F\big(E(\tilde{y}_{i*}), \boldsymbol{\Psi}\big),$$

with $\widehat{\boldsymbol{\Psi}}$ representing the set of the estimated nuisance parameters and

$$E(\tilde{y}_{i*}) = g\big(\tilde{\beta}_{0,i*} + \tilde{\mu}_{N,i*}\tilde{T}_{i*} + \tilde{\boldsymbol{\beta}}_{i*}^{\boldsymbol{PF}}\boldsymbol{x}_{i*}^{\boldsymbol{PF}} + \tilde{\boldsymbol{\beta}}_{\boldsymbol{N},i*}^{\boldsymbol{EM}}\boldsymbol{x}_{i*}^{\boldsymbol{EM}}\tilde{T}_{i*}\big), \qquad (3)$$

according to Equation (1) and its explanation.

Variable selection, model validation and software

To avoid overfitting, Equation (1) requires variable selection among all possible effect modifiers and prognostic factors and Equation (2) among all possible treatment predictors. We combined expert advice with the Least Absolute Shrinkage and Selection Operator (LASSO) approach to select variables (see supplemental Appendix S2) [23]. We did not extend the shrinkage to parameter estimation: model fitting was done using conventional least-squares and maximum-likelihood methods. Of note, a literature review could also help in identifying variables to include in the model. We investigated the validity of the prediction framework in development and validation samples. In particular, we assessed the treatment prediction model using receiver operating characteristic (ROC) and calibration curves. As a rule of thumb, an area under the ROC curve above 0.7 indicates moderate, and an area above 0.9 high accuracy [24]. Predicted outcomes in validation samples were compared graphically to observed outcomes. All analyses were done in R (https://www.r-project.org/). The R packages and code used are described in Appendix S5.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Results

Predicting effectiveness in the real world

In our case study, N was tocilizumab (TCZ) combined with cDMARDs, and the comparator intervention C was a cDMARD treatment. Information on the efficacy of TCZ was obtained from the TOWARD trial [16]. The SCQM registry (see supplemental Box S1) provided data on the socio-demographic and clinical characteristics of patients treated with different RA drugs in Switzerland. In discussion with two expert rheumatologists (A.F., S. R.) we first identified rituximab (RTX) as a drug that is comparable with TCZ. Patients receiving RTX would also likely receive TCZ: both drugs are biologics, typically administered after failure of a first anti-TNF agent. We therefore defined S as RTX in combination with any cDMARD(s). Based on the expert advice and the literature (e.g. [25]), we categorized covariates into potential prognostic factors, effect modifiers and treatment predictors. Table 1 summarizes the characteristics of patients on the relevant treatments in the TOWARD and REFLEX trials [16,17] and in the Swiss and British RA registries [18,19]. Compared to the registries the patients enrolled in the RCTs appeared to be younger, more likely to be female, more likely to have greater disease activity (as indexed by higher DAS28 scores), and more likely to be on steroids.

Next, we parameterized Equation (1). We centered all continuous covariates. For interpretability reasons, we did not center our integer covariates, i.e. the number of previous and the number of concomitant medications. Coefficients for the intercept and prognostic factors, based on the trial data on N and C, and the registry data on S and C, are presented in the upper part of Table 2. For example, the predicted 6-month change in DAS28 is -1.295 in a hypothetical patient taking C, whose rheumatoid factor (RF) is negative, who had never been exposed to any anti-TNF treatment, and whose other characteristics all correspond to the overall population means. It would be $-1.295 + 0.369 + 2 * 0.266 = -0.394$ if the patient was RF-positive and characterized by two previous anti-TNF treatments. Of note, a higher baseline DAS28 increases the expected decrease in DAS28 at 6 months. The bottom part of Table 2 shows the coefficients for the relative treatment effect of N versus C and the influence of the effect modifiers, estimated from the RCT data. For example, the predicted DAS28 score would be 1.078 lower in a patient taking N than in a patient taking C, assuming their baseline characteristics were identical, they both were RF-negative, and they had never taken any anti-TNF agents. Note that RF-positivity and exposure to previous anti-TNF medications would increase the difference between the effects of N and C.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

We used Equation (2) to define the profile of patients who will likely receive N after its launch and then assessed the effectiveness of N in a real-world routine setting. First, we simulated a real-world patient population from the Swiss SCQM registry, using the patient characteristics observed in patients on S or C after 2005, after removal of duplicate records. We generated 10,000 subjects. Their characteristics are shown in supplemental Table S1. The roughly 40% of patients assigned to receive N had a higher DAS28 score, a longer disease duration, were more likely to be RF-positive and on steroids, and had greater exposure to anti-TNF drugs than the roughly 60% of patients constituting the comparator group (C). We then used Equation (3) to predict treatment responses: Figure 1 illustrates the predicted effectiveness of N versus C determined as the gap between treatment outcomes measured in registry patients taking C and treatment outcomes predicted in simulated individuals assigned to receive N. The mean change in DAS28 at 6 months was -2.101 (standard deviation (SD): 1.494) in the simulated patients receiving N and -0.792 (SD: 1.499) in the observed patients treated with C.

Validation

We compared predicted with observed treatment outcome. For this purpose, we analyzed observational data on $\boldsymbol{N}$ (TCZ and any cDMARD(s)) and $\boldsymbol{C}$ (any cDMARD(s)) from the Swiss registry. Figure 2 shows that the framework predicted the 6-months changes in DAS28 quite accurately in both treatment groups. For $\boldsymbol{C}$, the mean changes in DAS28 were $\boldsymbol{-0.792}$ (observed, SD $\boldsymbol{1.499}$), $\boldsymbol{-0.430}$ (simulated population likely to receive $\boldsymbol{C}$, SD $\boldsymbol{1.413}$) and $\boldsymbol{-0.972}$ (observed in TOWARD trial, SD $\boldsymbol{1.205}$). For $\boldsymbol{N}$, the corresponding mean changes in the DAS28 were $\boldsymbol{-1.873}$ (SD $\boldsymbol{1.220}$), $\boldsymbol{-2.101}$ (SD $\boldsymbol{1.494}$) and $\boldsymbol{-2.914}$ (SD $\boldsymbol{1.416}$), respectively. The slight overestimation of treatment success in the $\boldsymbol{N}$ group and the underestimation in the $\boldsymbol{C}$ group could be explained by residual confounding. We also studied the predictive performance of the treatment assignment model described by Equation (2). The model discriminated well between patients receiving $\boldsymbol{N}$ and patients receiving $\boldsymbol{C}$, with an area under the ROC curve of 0.91. Finally, we assessed transferability across countries (external, geographical validity [26]) by developing the prediction approach using data from the Swiss registry and making predictions for patients from the British registry. We found that the accuracy of predicting treatment was poor (area under the ROC curve 0.35). The predicted and observed changes in DAS28 for patients on $\boldsymbol{N}$, based on the modelling framework developed on patients from the Swiss registry, were however fairly accurate: -2.325 and -2.587, respectively. In a further analysis, we trained the model using the British registry data and made predictions for Swiss RA patients. Results from this analysis were similar: prediction of treatment was relatively poor (area under the ROC curve 0.66) but predicted and observed treatment outcomes for $\boldsymbol{N}$ were again similar (-2.250 and -1.873, respectively). Details on the validation studies are presented in supplemental Appendix S3.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Discussion

We developed a modelling approach to predict the effectiveness of a new drug, assuming that evidence on its efficacy was available from RCTs at the time of the analysis, but no observational data on its effectiveness in the real world. The prediction process comprises two stages: firstly, a typical sample of real-world patients who are likely to receive the new treatment is identified. Secondly, the treatment and likely treatment outcomes are predicted for these patients. Both stages account for the prevalence of all relevant patient and disease characteristics in the target patient population. The modelling framework considers different sources of evidence, including expert advice. It is a compromise between purely statistical and purely qualitative approaches to the prediction of drug effectiveness, enriching conclusions from RCTs with insights from observational data and everyday clinical practice. The statistical methods employed, generalized linear models, are available in many software packages and well documented in the literature [27]. Applied to a case study in rheumatoid arthritis, the modelling approach accurately predicted the effectiveness of a new biologic intervention.

The suitability of the proposed modelling approach relies on three key requisites: firstly, the target population for the new drug must resemble an observable patient population receiving an approved drug in daily clinical practice. The existence of such a drug is not guaranteed, and its identification requires in-depth consultations with clinical experts. Guidelines that cover the new drug may be available and detail the characteristics of the patients who should receive the drug. Secondly, both individual participant level RCT data on the new drug and individual participant level observational data on the existing, approved drug must be available and include information on relevant covariates. Observational evidence must also be available for the control treatments included in the RCT. Thirdly, the methods of data collection and reporting must be comparable between the RCTs and observational studies.

Our approach has several limitations. Whether the requirements outlined above are fulfilled can only be answered on a case-by-case basis. It is also impossible to judge upfront whether the available databases provide evidence on all relevant effect-modifiers. We may miss important information if considering only the effect modifiers that were considered by the authors of published RCTs. Non-adherence to medication, for instance, may be an important effect modifier in many real-world patient populations. In general, any unmeasured or undetected confounder may lead to biased predictions. Also, it is important that the observational databases are of high quality, informative and representative of the general patient population. A variable might not have been measured or reported the same way across the different data sources. In this case, definitions and reporting

should be standardized as much as possible so that they are consistent between observational databases and RCTs. Missing values may be imputed if the amount of missing is limited. The methods of data collection should be well documented for all data sources. The proposed prediction framework should not be applied if the overlap between the RCT and the target real-world population is small, i.e. if the RCT had narrow inclusion criteria, for example by including middle-aged men only. The more pragmatic the trial design is the more reliable predictions will likely be for a real-world population.

The model can be used in several decision making scenarios. For example, the model could be used early in drug development to optimize the design of RCTs by modelling different scenarios of effectiveness, based on different assumptions for efficacy. Also, although not addressed by the GetReal project, the modeling framework could be used to predict safety-related treatment outcomes, and to inform cost-effectiveness analyses. The use of modelling and simulation is becoming more widespread in drug development [28] with the results increasingly being used to support regulatory submissions. To facilitate the uptake of modelling in decision-making, the methods need to be completely transparent. A range of stakeholders can then understand the methods employed, the key assumptions made, the quality of the data, as well as potential biases and limitations of the analyses.

In contrast to previous studies [7], our work carefully addresses the validation of the model and illustrates possible solutions in applications to our case study. Internal model validation showed good accuracy for both prediction stages. When investigating the external, geographical validity of our model, we found that the accuracy of predicted treatments was poor, probably reflecting the large differences between the Swiss and British healthcare systems. This may be due to country-specific differences in treatment guidelines, treatment costs, and reimbursement policies. Pharmaceutical researchers and policy makers should thus be aware that predictions may be inappropriate for healthcare systems other than the one from which the observational data used for model development originate. The model predicting treatment outcome may still be useful and accurate, for example to assess the effectiveness of the new drug (independent of treatment decisions) in a subgroup of patients.

In general, to avoid missing valuable information, we suggest considering a broad range of data sources. To flexibly weigh and integrate different sources of aggregate and individual-level data, the purely frequentist inference concept described in this work may be translated to a Bayesian setting [27]. The complexity of relationships between the variables used in predicting outcomes may vary and more complex correlations structures may be required to fit the nature of the problem. Furthermore, results from (network) meta-analyses

[29,30] may be considered to assess the comparative effectiveness of TCZ versus other biologic DMARDs. A long-term view on treatment effect and outcome may sometimes be desirable. Dynamic treatment regimens with time-varying confounders and censoring information should then be considered, as discussed by Hernan and Robins in their forthcoming book [31]. If more than two treatment arms are investigated, e.g. to examine the dose-specific efficacy of a new drug, an appropriately implemented multinomial model is needed to predict treatment decision. These are just a few of many effectiveness questions that may be addressed by suitable extensions of our multi-stage model.

Conclusion

We developed a novel modelling tool to predict treatment effectiveness prior to launch of a new drug, i.e. when only Phase II or III clinical trials are available, and illustrated its application in a case study. Building on its intuitive structure, we envisage further methodological developments to expand its application to a wider range of drug effectiveness problems.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

References

[1] S. R. Cole and E. A. Stuart, "Generalizing Evidence From Randomized Clinical Trials to Target Populations The ACTG 320 Trial," *American Journal of Epidemiology*, vol. 172, no. 1, pp. 107–115, Jul. 2010.

[2] M. F. Drummond *et al.*, "Key principles for the improved conduct of health technology assessments for resource allocation decisions," *International Journal of Technology Assessment in Health Care*, vol. 24, no. 03, Jul. 2008.

[3] C. Nordon *et al.*, "The "Efficacy-Effectiveness Gap": Historical Background and Current Conceptualization.," *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, vol. 19, no. 1, pp. 75–81, Jan. 2016.

[4] H.-G. Eichler *et al.*, "Bridging the efficacy–effectiveness gap: a regulator's perspective on addressing variability of drug response," *Nature Reviews Drug Discovery*, vol. 10, no. 7, pp. 495–506, Jul. 2011.

[5] M. R. Stratton *et al.*, "Genomics and the Continuum of Cancer Care," *New England Journal of Medicine*, vol. 364, no. 4, pp. 340–350, Jan. 2011.

[6] R. L. Tannen, M. G. Weiner, and D. Xie, "Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings.," *BMJ (Clinical research ed.)*, vol. 338, p. b81, Jan. 2009.

[7] K. Panayidou *et al.*, "GetReal in mathematical modelling: a review of studies predicting drug effectiveness in the real world," *Research Synthesis Methods*, vol. epub ahead, pp. 1–14, 2016.

[8] W. C. Levy, "The Seattle Heart Failure Model: Prediction of Survival in Heart Failure," *Circulation*, vol. 113, no. 11, pp. 1424–1433, Mar. 2006.

[9] P. Ponikowski *et al.*, "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC," *European Heart Journal*, vol. 37, no. 27, pp. 2129–2200, Jul. 2016.

[10] C. W. Yancy *et al.*, "2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines," *Circulation*, vol. 128, no. 16, pp. e240–e327, Oct. 2013.

[11] GetReal consortium, "IMI GetReal - Real-Life Data In Drug Development." [Online]. Available: http://www.imi-getreal.eu/. [Accessed: 13-Feb-2016].

[12] M. Egger, K. G. M. Moons, C. Fletcher, and GetReal Workpackage 4, "GetReal: from efficacy in clinical trials to relative effectiveness in the real world.," *Research synthesis methods*, vol. 7, pp. 278–281, Jul. 2016.

[13] K. D. Pile, G. G. Graham, and S. M. Mahler, "Disease-Modifying Anti-Rheumatic Drugs," in *Encyclopedia of Inflammatory Diseases*, M. Parnham, Ed. Basel: Springer Basel, 2015, pp. 1–13.

[14] M. C. Genovese *et al.*, "Interleukin-6 receptor inhibition with tocilizumab reduces disease activity in rheumatoid arthritis with inadequate response to disease-modifying antirheumatic drugs: The tocilizumab in combination with traditional disease-modifying antirheumatic drug the," *Arthritis & Rheumatism*, vol. 58, no. 10, pp. 2968–2980, Oct. 2008.

[15] S. B. Cohen *et al.*, "Rituximab for rheumatoid arthritis refractory to anti–tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks," *Arthritis & Rheumatism*, vol. 54, no. 9, pp. 2793–2806, Sep. 2006.

[16] T. Langenegger, J. Fransen, A. Forster, M. Seitz, and B. A. Michel, "Klinisches Qualitätsmanagement bei der Rheumatoiden Arthritis," *Zeitschrift für Rheumatologie*, vol. 60, no. 5, pp. 333–341, Oct. 2001.

[17] E. M. Dennison, J. Packham, and K. Hyrich, "The BSRBR-RA at 15 years: Providing real-world insight into the effectiveness and safety of biologic therapies," *Rheumatology*, p. kew053, Mar. 2016.

[18] J. Fransen, G. Stucki, and P. L. C. M. van Riel, "Rheumatoid arthritis measures: Disease Activity Score (DAS), Disease Activity Score-28 (DAS28), Rapid Assessment of Disease Activity in Rheumatology (RADAR), and Rheumatoid Arthritis Disease Activity Index (RADAI)," *Arthritis & Rheumatism*, vol. 49, no. S5, pp. S214–S224, Oct. 2003.

[19] C. Gabay *et al.*, "Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): a randomised, double-blind, controlled phase 4 trial," *The Lancet*, vol. 381, no. 9877, pp. 1541–1550, May 2013.

[20] S. J. Tannenbaum, N. H. G. Holford, H. Lee, C. C. Peck, and D. R. Mould, "Simulation of correlated continuous and categorical variables using a single multivariate distribution," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 33, no. 6, pp. 773–794, Dec. 2006.

[21] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective: Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 3, pp. 273–282, Jun. 2011.

[22] Swets, John A., "Measuring the Accuracy of Diagnostic Systems," *Science*, vol. 240, no. 4857, pp. 1285–1293, 1988.

[23] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, vol. 33, no. 1, 2010.

[24] X. Robin *et al.*, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.

[25] C. S. Crowson, E. L. Matteson, J. M. Davis, and S. E. Gabriel, "Contribution of obesity to the rise in incidence of rheumatoid arthritis," *Arthritis Care & Research*, vol. 65, no. 1, pp. 71–77, Jan. 2013.

[26] P. C. Austin, D. van Klaveren, Y. Vergouwe, D. Nieboer, D. S. Lee, and E. W. Steyerberg, "Geographic and temporal validity of prediction models: different approaches were useful to examine model performance," *Journal of Clinical Epidemiology*, Jun. 2016.

[27] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, 3. ed. Boca Raton, Fla.: CRC Press/Chapman & Hall, 2014.

[28] EFPIA MID3 Workgroup *et al.*, "Good Practices in Model-Informed Drug Discovery and Development: Practice, Application, and Documentation: Good Practices in Model-Informed Drug Discovery and Development," *CPT: Pharmacometrics & Systems Pharmacology*, vol. 5, no. 3, pp. 93–122, Mar. 2016.

[29] T. P. A. Debray *et al.*, "GetReal in Individual Participant Data (IPD) Meta-Analysis: A review of the methodology," *Journal of Research Synthesis Methods*, vol. 6, no. 4, pp. 293–309, 2015.

[30] O. Efthimiou *et al.*, "GetReal in network meta-analysis: a review of the methodology," *Research Synthesis Methods*, vol. in press, no. November 2014, 2016.

[31] M. A. Hernan and J. M. Robins, *Causal Inference*. Boca Raton, Fla.; London: CRC ; Taylor & Francis [distributor], 2017.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Box 1: Notation

i	Person index in a patient population used for model development and estimation	
i^*	Person index in a new patient population used for making predictions	
j	Treatment classifier to distinguish between a new drug (N) and an existing drug (S) which is similar to N in terms of target patient and disease characteristics; $j \in \{N, S\}$	
Variables		
y_i	Treatment outcome in patient i	
$\boldsymbol{x_i^{PF}}$	Column vector of prognostic factors for patient i	
$\boldsymbol{x_i^{EM}}$	Column vector of effect modifiers for patient i	
$\boldsymbol{x_i^{TP}}$	Column vector of treatment predictors for patient i	
T_i	Treatment indicator for patient i; $T_i \coloneqq \begin{cases} 1 & \text{if patient i receives a certain treatment} \\ 0 & \text{if patient i receives a comparator treatment} \end{cases}$	
D_i	Study indicator for patient i; $D_i \coloneqq \begin{cases} 1 & \text{if patient } i \text{ was an RCT participant} \\ 0 & \text{if patient } i \text{ was treated in clinical routine} \end{cases}$	
w	(Optional) Relative weight of RCT vs. real-world evidence	
Model describing/predicting treatment outcome		
Parameter set $\boldsymbol{\Theta_I} \coloneqq \{\beta_0, \mu_j, \boldsymbol{\beta^{PF}}, \boldsymbol{\beta_j^{EM}}\}$		**Data sources**
β_0	Intercept term	RCT and OBS
μ_j	Relative effect of treatment *j* vs. a comparator intervention	RCT or OBS
$\boldsymbol{\beta^{PF}}$	Row vector of the effects of the prognostic factors on disease status	RCT and OBS
$\boldsymbol{\beta_j^{EM}}$	Row vector of the effect-modifying effects of treatment *j* vs. a comparator drug	RCT or OBS
$\boldsymbol{\Psi}$	Set of nuisance parameters	RCT and OBS
$\boldsymbol{\sigma^2}$	Variance parameter in a Gaussian setting	RCT and OBS
Model describing/predicting treatment assignment		
Parameter set $\boldsymbol{\Theta_{II}} \coloneqq \{\gamma_0, \boldsymbol{\gamma}\}$		**Data sources**
γ_0	Intercept term	OBS
$\boldsymbol{\gamma}$	Row vector of the effects of the treatment predictors on the decision "*j* vs. comparator"	OBS

PF, prognostic factor; EM, effect modifier; TP, treatment predictor; RCT, randomized controlled trial; OBS, observational data.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Table 1: Baseline characteristics of patients enrolled in the randomized controlled trials and routine databases.

	Randomized controlled trials				Observational data					
	TOWARD		REFLEX		SCQM			BSRBR-RA		
	TCZ+cDMARD (N)	***cDMARD (C)***	***RTX+cDMARD (S)***	***cDMARD (C)***	***TCZ+cDMARD (N)***	***RTX+cDMARD (S)***	***cDMARD (C)***	***TCZ+cDMARD (N)***	***RTX+cDMARD (S)***	***cDMARD (C)***
N	803	413	308	209	265	290	895	259	629	1137
Effect modifiers										
No. of previous anti-TNF agents	0.4 (0.7)	0.4 (0.7)	1.5 (0.7)	1.5 (0.7)	1.2 (0.9)	1.1 (0.9)	0.1 (0.3)	1.0 (0.9)	1.1 (0.8)	0.3 (0.6)
RF-positivity[a,b]	78%	75%	77%	79%	72%	83%	74%	62%	69%	63%
Prognostic factors										
DAS28 score	6.4 (1.0)	6.3 (0.9)	6.9 (1.0)	6.8 (0.9)	4.3 (1.2)	4.6 (1.3)	4.2 (1.6)	5.5 (1.3)	5.7 (1.1)	4.9 (1.2)
Disease duration	9.8 (9.1)	9.8 (8.8)	12. 8 (8.3)	11.7 (7.7)	9.6 (9.4)	10.8 (9.0)	6.1 (8.3)	11.5 (9.6)	13.6 (9.7)	10.7 (10.1)
Body-mass index	27.8 (6.3)	27.5 (6.3)	28.3 (6.9)	29.5 (7.3)	25.9 (4.9)	26.3 (4.9)	25.2 (4.8)	29.5 (7.1)	28.6 (7.7)	27.2 (6.3)
Treatment predictors										
No. of previous cDMARD	1.2 (1.3)	1.3 (1.3)	2.6 (1.8)	2.4 (1.8)	1.9 (0.9)	1.8 (1.0)	0.2 (0.6)	1.9 (1.3)	2.6 (1.6)	1.8 (1.8)
No. of current cDMARD	1.3 (0.6)	1.3 (0.7)	1.0 (0.0)	1.0 (0.0)	1.1 (0.3)	1.1 (0.3)	1.2 (0.5)	1.3 (0.6)	1.2 (0.5)	1.4 (0.6)
On steroid treatment	51%	55%	72 %	71 %	50%	45%	23%	34 %	39%	30%
Other variables										
Age	53.0 (12.6)	53.5 (13.1)	52.3 (12.3)	52.8 (12.6)	59.1 (9.9)	56.8 (11.7)	55.5 (14.0)	55.9 (12.8)	58.6 (11.7)	60.5 (12.2)
Sex (female)	81%	84%	81%	82%	76%	77%	74%	78%	78%	75%
Smoking	17%	17%	N/A	N/A	20%	33%	63%	28%	22%	23%

Mean (standard deviation) or total percentage are shown. TCZ, tocilizumab; RTX, rituximab; cDMARD, conventional Disease Modifying Anti-Rheumatic Drug; TOWARD, Tocilizumab in Combination With Traditional DMARDs trial; REFLEX, Randomized Evaluation of Long-Term Efficacy of Rituximab trial; SCQM, Swiss Clinical Quality Management in rheumatic diseases; BSRBR-RA, British Society for Rheumatology Biologics Registry - Rheumatoid Arthritis; TNF, tumor necrosis factor; RF, rheumatoid factor; DAS, disease activity score.
See text for definition of N, S, C.
a, also a prognostic factor; *b*, also a treatment predictor.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Table 2: Coefficients for the model intercept and the prognostic factors - estimated from the TOWARD data on cDMARDs (C) and a combined TCZ-cDMARDs treatment (N), and from the SCQM registry data on C and a combined RTX-cDMARDs treatment (S) - , and coefficients for the main treatment effect of N versus C and the effect modifiers (estimated from the TOWARD data alone), with 95% confidence intervals.

Intercept	$\hat{\beta}_0$	95% confidence interval
	-1.295	-1.419 , -1.172
Prognostic factors	$\widehat{\boldsymbol{\beta}}^{\boldsymbol{PF}}$	95% confidence interval
RF-positivity*	0.369	0.167, 0.572
Baseline DAS28 score	-0.363	-0.410, -0.316
Disease duration*	0.004	-0.003, 0.012
Body mass index*	0.016	0.005, 0.027
No. of previous anti-TNF agents*	0.266	0.088, 0.443
Relative treatment effect	$\hat{\mu}_N$	95% confidence interval
N versus C	-1.078	-1.360, -0.796
Effect modifiers	$\widehat{\boldsymbol{\beta}}^{\boldsymbol{EM}}$	95% confidence interval
RF-positivity*	-0.690	-0.996, -0.384
No. of previous anti-TNF agents*	-0.056	-0.277, 0.164

*Variable included based on expert advice

TCZ, tocilizumab; RTX, rituximab; cDMARD, conventional Disease Modifying Anti-Rheumatic Drug; TOWARD, Tocilizumab in Combination With Traditional DMARDs trial; SCQM, Swiss Clinical Quality Management in rheumatic diseases; RF, rheumatoid factor; DAS, disease activity score; TNF, tumor necrosis factor

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Table 3: Coefficients for the treatment predictors – estimated from the SCQM registry data on cDMARDs ($\boldsymbol{C}$) and a combined RTX-cDMARDs treatment ($\boldsymbol{S}$) –, with 95% confidence intervals.

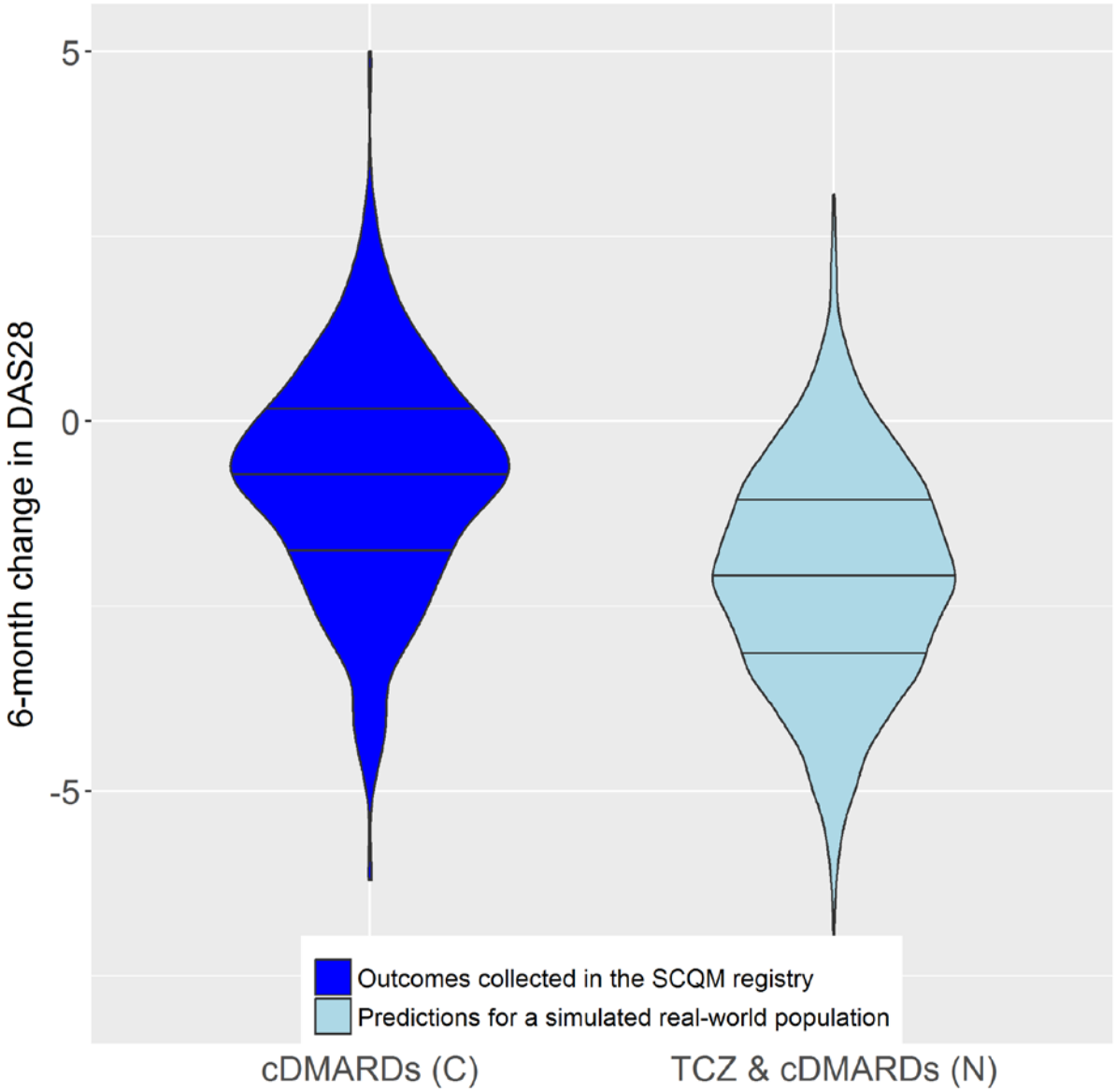
Intercept	$\hat{\gamma}_0$	95% confidence interval
	-3.222	-5.053, -1.501
Treatment predictors	$\hat{\boldsymbol{\gamma}}$	95%-CI
RF-positivity*	1.369	0.457, 2.370
Baseline DAS28	0.351	-0.092, 0.624
Disease duration*	0.041	0.000, 0.083
No. of previous cDMARDs	1.220	0.764, 1.783
No. of previous anti-TNF agents	1.456	0.831, 2.145
No. of concomitant cDMARDs	-2.189	-3.317, -1.187
On steroids	0.726	0.092, 1.511

*Variable included based on expert advice

RTX, rituximab; cDMARD, conventional Disease Modifying Anti-Rheumatic Drug; SCQM, Swiss Clinical Quality Management in rheumatic diseases; RF, rheumatoid factor; DAS, disease activity score; TNF, tumor necrosis factor

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Figure 1: Observed effectiveness of cDMARDs ($\boldsymbol{C}$) and predicted effectiveness of a combined TCZ-cDMARDs treatment ($\boldsymbol{N}$).

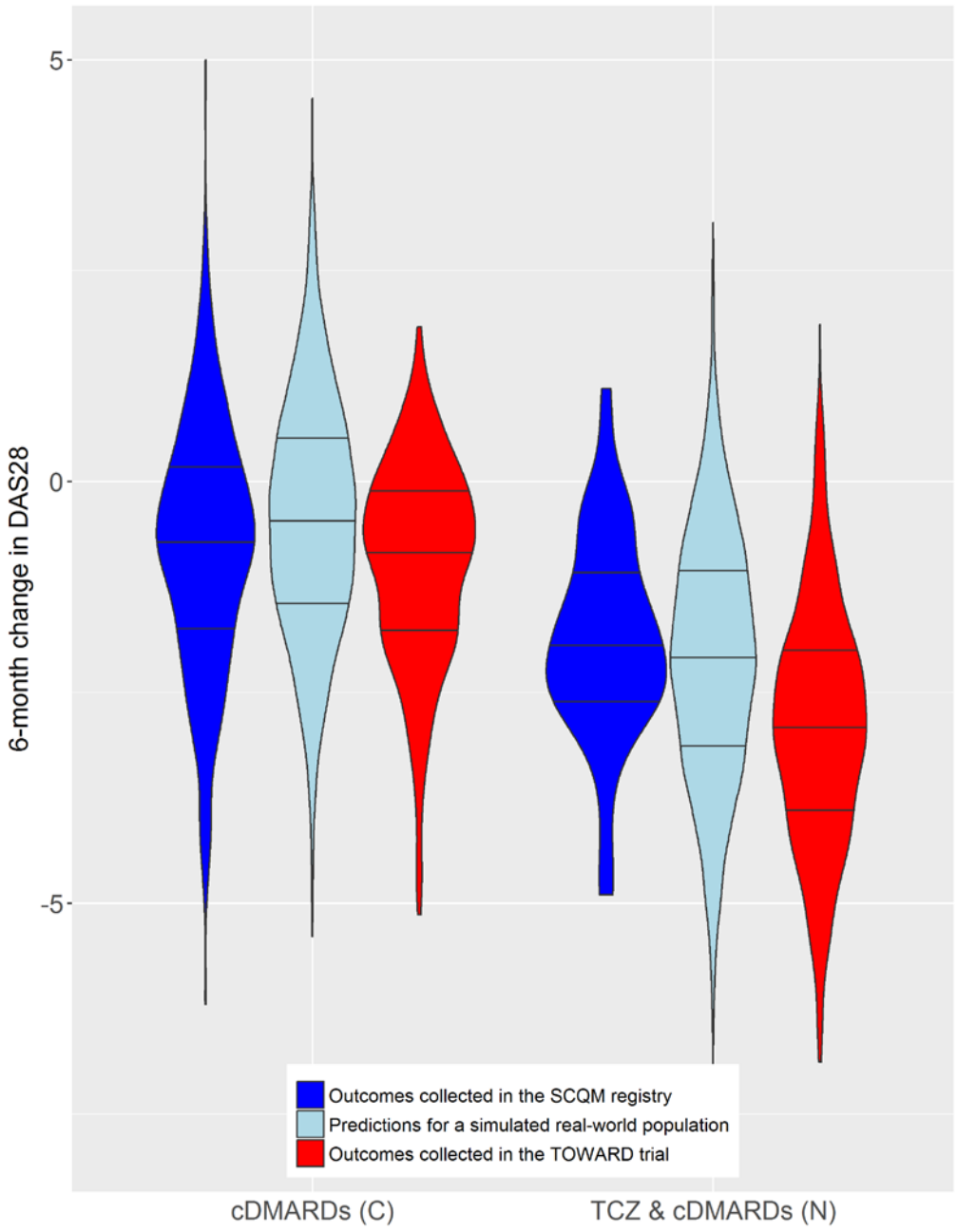


TCZ, tocilizumab; cDMARD, conventional Disease Modifying Anti-Rheumatic Drug; SCQM, Swiss Clinical Quality Management in rheumatic diseases; DAS, disease activity score

The violin plots show the medians (middle horizontal lines) and quartiles (lower and upper horizontal lines) of the 6 months change in DAS28. The means (standard deviations) of the changes were $\boldsymbol{-0.792}$ $(\boldsymbol{1.499})$ in the observed patients on treatment $\boldsymbol{C}$ and $\boldsymbol{-2.093}$ $(\boldsymbol{1.483})$ in the simulated patients on treatment $\boldsymbol{N}$.

Published in final edited form as : *Med Decis Making*. 2018 Aug;38(6):719-729. doi: 10.1177/0272989X18775975

Figure 2: Clinical performance of cDMARDs (C) and a combined TCZ-cDMARDs treatment (N).



TCZ, tocilizumab; cDMARD, conventional Disease Modifying Anti-Rheumatic Drug; SCQM, Swiss Clinical Quality Management in rheumatic diseases; TOWARD, Tocilizumab in Combination With Traditional DMARDs trial; DAS, disease activity score

The violin plots show the medians (middle horizontal lines) and quartiles (lower and upper horizontal lines) of the changes at 6 months in the DAS28 score. The means (standard deviations) of the changes were $-0.792\ (1.499)$ /$-0.455\ (1.424)$ /$-0.972\ (1.205\ )$ in the observed/simulated/RCT patients on treatment C, and $-1.873\ (1.220)$ / $-2.093\ (1.483)$ /$-2.914\ (1.416)$ in the observed/simulated/RCT patients on treatment N.